Gouvernement du Canada
Ministère des Communications

Government of Canada
Department of Communications

# Le Centre canadien de recherche sur l'informatisation du travail

# Canadian Workplace Automation Research Centre

Canada

# BILAN ET PERSPECTIVES DE LA TRADUCTION ASSISTEE PAR ORDINATEUR

Pierre Isabelle

Centre canadian de recherche sur
l'informatisation du travail (CCRIT)

Copyright Ministre des Approvisionnements et Services Canada 1988.

No de cat. Co 28-1/19-1988F

ISBN 0-662-16397-4

Les opinions émises dans ce rapport n'engagent que l'auteur.

* This report is also available in English.

Le présent exposé a pour objectif de présenter le programme de recherche et développement du CCRIT en traduction assistée par ordinateur, en situant ce programme dans la perspective de notre appréciation de l'état actuel des travaux en matière d'automatisation de la traduction.

## LE PROBLÈME DE L'AUTOMATISATION DE LA TRADUCTION

Dès l'avènement des ordinateurs numériques, dans les années cinquante, il se trouva beaucoup de gens pour croire à l'apparition imminente de la machine à traduire universelle. Pourtant, un examen sommaire de la situation actuelle met en évidence le fait que trente années d'efforts ont produit bien peu de résultats. La plupart des traducteurs d'aujourd'hui travaillent toujours avec des méthodes très similaires à celles d'il y a cent ans.

Au fil des ans, l'optimisme initial envers la traduction automatique a fait place à une grande méfiance, sinon au pessimisme le plus absolu dans les milieux informés. On s'est rendu compte que la traduction entre langues naturelles ne pouvait en aucun cas être assimilée à un processus primaire de transcodage.

Il est vite devenu apparent que l'information explicitement présente dans un texte quelconque laisse subsister une grande part d'indétermination quant au contenu informationnel de ce texte. Dès 1960, Bar-Hillel (qui avait pourtant été auparavant très optimiste quant à l'avenir de la traduction automatique) avançait la thèse de l'impossibilité de la traduction entièrement automatique avec résultats de haute qualité. Bar-Hillel fondait sa démonstration[1] sur des exemples comme le suivant :

(1) The box is in the pen.

Un dictionnaire bilingue nous dira que pen se traduit tantôt par plume, tantôt par parc (de bébé). Le choix correct dans la traduction de l'exemple de Bar-Hillel doit reposer non sur des renseignements explicites dans le texte, mais plutôt sur notre connaissance du monde (par exemple, la taille relative des objets, la probabilité relative de différentes situations, etc.). Il est très facile de fournir d'autres exemples dont la traduction correcte suppose l'accès à notre connaissance de la réalité extralinguistique :

(2) ... federal and provincial governments ...
(3) John dropped the glass on the table. It broke.
(4) Discard used filter and replace it.

Dans le cas de l'exemple 2 le traducteur devra faire un choix quant à l'accord en nombre des adjectifs; le choix de "fédéral et provinciaux" résulte de sa connaissance d'une certaine réalité politique. Pour traduire correctement "it" ("il" ou "elle") en 3 il faut déterminer ce qui s'est cassé (le verre ou la table); c'est le sens commun qui, à défaut d'indications contraires, nous fait préférer la traduction par "il". Enfin, la traduction de l'exemple 4 force le traducteur à choisir entre "remettre en place" et "remplacer" comme traduction de "replace". Encore ici, on fait appel au sens commun : lorsqu'on jette un objet à la poubelle ce n'est pas normalement avec l'intention de le reprendre et de le remettre en place.

La conclusion inévitable de telles constatations, c'est que la traduction entièrement automatique de bonne qualité de textes arbitrairement choisis n'est possible que dans la mesure où nous disposons de machines intelligentes, c'est-à-dire de machines capables d'accéder non seulement au contenu linguistique explicite des textes (morphologie, syntaxe, sémantique), mais aussi à leur contenu implicite fondé sur une organisation cognitive globale (sens commun, savoir encyclopédique, raisonnements, etc.).

Si au cours des 15 dernières années, l'entreprise ambitieuse connue sous le nom d'intelligence artificielle a enregistré des progrès tangibles, elle nous a surtout permis de mieux saisir la complexité formidable des mécanismes cognitifs humains. Nous sommes encore bien loin de la machine à penser universelle. Par conséquent, nous sommes encore loin de la machine à traduire universelle.

## LA RECHERCHE DE SOLUTIONS SOUS-OPTIMALES

Le marché de la traduction connaît depuis plusieurs années, à l'échelle mondiale, une croissance extrêmement soutenue. Les enjeux économiques et politiques qui alimentent ce marché ne cessent de prendre de l'importance.

À défaut d'une solution optimale au problème de la traduction (la machine à traduire universelle), il importe de trouver des palliatifs qui aideront les traducteurs à faire face à une demande de plus en plus difficile à satisfaire et ce, à un coût acceptable. Ces palliatifs se fonderont sur des compromis concernant l'un ou plusieurs des attributs suivants de la machine à traduire universelle : a) capacité de produire des traductions de haute qualité; b) automatisation totale; c) capacité de traduire n'importe quel texte.

### (A) Les traductions approximatives

Il arrive parfois que des spécialistes doivent procéder à l'"écrémage" de masses de documentations en langue étrangère pour y dépister un petit sous-ensemble intéressant. Des machines capables de fournir de grossières ébauches peuvent jouer un rôle utile, quitte à ce que les documents jugés intéressants soient ensuite traduits manuellement. Les situations où ce genre de technique est utilisable (p. ex., l'intelligence militaire) sont toutefois l'exception plutôt que la règle.

### (B) La traduction semi-automatique

Un ensemble de techniques d'applicabilité plus générale se fonde sur une automatisation partielle du processus de traduction.

Dans le cas de la traduction humaine assistée, le traducteur humain conserve l'initiative complète du processus de traduction, mais il a recours à une machine pour automatiser certaines tâches périphériques. Parmi les outils potentiellement utiles au traducteur, on peut inclure la télécommunication de textes, l'accès à des bases de données lexicographiques et terminologiques, l'accès à des banques de textes traduits, le traitement

de texte multilingue, les vérificateurs d'orthographe, etc. Cette approche modeste sur le plan technologique offre l'avantage d'être applicable dans la plupart des contextes de traduction. On ne peut que déplorer, faisant écho à Kay[2] qu'elle demeure jusqu'à maintenant fort peu exploitée.

La traduction machine assistée par l'humain se distingue de l'approche précédente en ce qu'elle confie à la machine l'initiative principale du processus de traduction, mais lui fournit une assistance humaine pour résoudre les problèmes trop difficiles. L'intervention humaine peut être effectuée avant le processus de traduction machine par des annotations ajoutées au texte; on parle alors de systèmes fondés sur une pré-édition humaine (cf. le système CULT de l'Université de Hong-Kong). L'intervention peut également prendre la forme d'indications fournies à la machine sur demande; en ce cas, il s'agit de systèmes dits interactifs (cf. le système ALPS). Finalement, l'intervention peut avoir lieu après coup, pour réparer les erreurs ou combler les vides, et on parle alors de post-édition. La plupart des systèmes de traduction automatisée ont recours à une certaine quantité de post-édition. Par contre, la pré-édition et l'interactivité posent des problèmes difficiles : l'information qu'exige la machine doit souvent être formulée dans des termes trop proches de la machine, trop loin de l'humain.

(C) La traduction automatique spécialisée

Si l'on veut retenir des exigences élevées de qualité et une automatisation relativement poussée, on peut alors renoncer aux systèmes "universels" et se tourner vers des systèmes conçus sur mesure pour des types très particuliers de textes. Dans cette catégorie, on trouve d'abord les systèmes à entrée contrôlée, c'est-à-dire des systèmes qui présupposent que le texte à traduire a été rédigé dans une langue simplifiée, suivant un ensemble rigide de conventions. Le système TITUS, qui traduit des résumés d'articles dans le domaine des textiles en est l'exemple le plus connu. Cette approche n'est applicable que lorsque les services de traduction ont un contrôle sur le processus de rédaction et que le sujet se prête bien à une langue extrêmement codifiée.

On peut aussi mettre au point des systèmes spécialisés pour des sous-langues naturelles[3]. Cette approche mise de l'avant par le groupe TAUM de l'Université de Montréal[4] se fonde sur l'existence de sous-systèmes linguistiques adoptés spontanément par des sous-communautés linguistiques dans certaines situations. Les prévisions météorologiques, les rapports d'activité boursière et les recettes de cuisine constituent autant d'exemples frappants de sous-langues naturelles. Dans chaque cas, les contraintes pragmatiques sur le contenu des textes s'accompagne de restrictions particulièrement fortes sur la langue qui véhicule ce contenu. Le système TAUM-MÉTÉO traduit quotidiennement les prévisions météorologiques d'Environnement Canada depuis 1977, et sa charge de travail approche maintenant 15 millions de mots par an. Le taux de post-édition requis est minime et ce système constitue sans contredit la réussite la plus spectaculaire à ce jour en traduction automatisée.

L'automatisation de la traduction peut donc à l'heure actuelle s'averer très rentable, pour autant que l'on sache mettre en oeuvre des "techniques de sous-optimisation" adéquates.

## INTÉRÊT RENOUVELÉ POUR LA TRADUCTION AUTOMATISÉE

L'optimisme initial quant à la possibilité de mettre au point la machine à traduire avait généré une activité intense de recherche et développement durant les années cinquante. La traduction automatique occupait alors une place prépondérante au sein des activités dans le domaine de la "mécanisation de la pensée", connu aujourd'hui sous le nom "d'intelligence artificielle".

En 1966, un comité mandaté par le gouvernement américain pour examiner les causes des échecs répétés enregistrés dans les tentatives de traduction automatique déposait un rapport maintenant connu sous le nom de "rapport ALPAC"[5]. Les auteurs y critiquaient sévèrement l'approche à court terme adoptée dans la plupart des projets et recommandaient que les crédits gouvernementaux soient réorientés vers la recherche à plus long terme en linguistique informatique et en intelligence artificielle. Ce rapport produisit alors un effet dévastateur sur le monde de la traduction automatisée, faisant disparaître la plupart des projets de R&D dans le domaine à travers le monde entier.

Bien qu'une lecture attentive du rapport indique qu'on y condamnait non pas la traduction automatisée en tant que champ d'étude mais plutôt les approches à courte vue en intelligence artificielle, il eut quand même pour effet de discréditer la notion même de traduction automatisée.

Encore aujourd'hui, l'activité de recherche en traduction automatisée est quasi inexistante aux États-Unis. Beaucoup de chercheurs en linguistique informatique préfèrent toujours concentrer leurs efforts sur des applications qui répondent à des besoins beaucoup moins évidents que ceux de la traduction.

Depuis 1980, on assiste toutefois à un regain d'intérêt croissant pour la traduction automatisée. Aux États-Unis, le mouvement est surtout perceptible dans le secteur privé. Trois nouveaux systèmes commerciaux (ALPS, WEIDNER et LOGOS) sont venus s'ajouter à SYSTRAN qui détenait un quasi-monopole depuis 1965, et ces systèmes semblent trouver de nombreux preneurs dans plusieurs pays. De plus, IBM et Digital Equipment poursuivent présentement des travaux en traduction assistée par ordinateur et en traduction automatique.

En Europe la CEE a lancé une activité de grande envergure lorsqu'elle entreprit l'ambitieux projet EUROTRA qui vise à construire un système capable de traduire entre les neuf langues de la communauté. Plusieurs pays européens financent leurs propres recherches universitaires (France, Hollande, Angleterre). Les firmes Siemens (Allemagne) et Philips (Hollande) sont à mettre au point leurs propres systèmes de traduction.

Mais c'est sans doute au Japon que l'activité est la plus intense. On y compte pas moins d'une quinzaine de projets de R&D dans les universités et les entreprises. De plus, la traduction automatique occupe une place centrale dans la seconde phase (présentement en cours de démarrage) du fameux projet sur les ordinateurs de cinquième génération.

## LE PROGRAMME DE RECHERCHE ET DÉVELOPPEMENT DU CCRIT

Le Centre canadien de recherche sur l'informatisation du travail est présentement à mettre sur pied un programme de recherche et développement en traduction assistée par ordinateur. Ce programme s'inspire des recommandations du rapport COGNOS[6] et reflète l'appréciation de l'état des travaux dont nous avons fait part dans les sections précédentes. Dans cette perspective, deux des trois volets du programme visent le développement de systèmes fondés sur des techniques précises de "sous-optimisation". Le troisième volet de notre programme a trait à la recherche orientée vers des systèmes de traduction plus puissants, basés sur des techniques relevant de l'intelligence artificielle.

### Volet I : Poste de travail spécialisé en traduction

Les seuls systèmes d'applicabilité générale qui soient présentement réalisables en traduction sont ceux qui laissent au traducteur humain l'entière initiative du processus de traduction, lui fournissant un ensemble intégré d'outils susceptibles de l'aider à accomplir sa tâche.

Le travail du traducteur est une tâche complexe, comportant beaucoup d'aspects différents. En conséquence, un poste de travail spécialisé en traduction devrait réunir et intégrer de façon adéquate une vaste gamme d'outils, parmi lesquels devraient figurer :

- la télécommunication de textes;
- les aides à la gestion des tâches de traduction;
- les aides à la recherche lexicographique et terminologique
  - un "cardex" informatisé
  - des dictionnaires naturels informatisés
  - l'accès à des banques terminologiques à distance
  - l'accès à des banques de textes traduits
  - des outils d'établissement de concordances
- les aides à la rédaction
  - un traitement de texte multilingue
  - des vérificateurs d'orthographe, de grammaire, etc.

Enfin, un tel poste de travail devrait prévoir l'intégration ultérieure de modules plus avancés, comme par exemple des systèmes de traduction automatique spécialisés.

### Volet II : Traduction automatique spécialisée

Nous avons vu plus haut que l'approche dite par sous-langue a permis d'obtenir des résultats spectaculaires dans le domaine des prévisions météorologiques (système TAUM-MÉTÉO). Il nous apparaît extrêmement important de mettre à profit ce genre de technique dans un nombre aussi grand que possible de domaines. Alors que le poste de travail prévu au volet I opte pour la généralité du champ d'application au détriment du degré d'automatisation, le présent volet effectue la démarche inverse. Les deux approches sont parfaitement complémentaires puisque, à moyen terme, le poste de travail fournira un environnement idéal pour l'implantation de modules de traduction automatique spécialisés.

Avant d'amorcer la mise au point de systèmes de traduction spécialisés, il est essentiel d'arriver à bien mesurer la complexité des sous-langues concernées. On s'attachera donc à inventorier les domaines d'application possibles et à évaluer leur complexité relative avec les meilleurs instruments de mesure possibles.

Ensuite seulement pourra-t-on lancer un ou plusieurs projets de développement, qui reposent sur l'utilisation de techniques et d'outils de programmation éprouvés en matière de traitement des langues naturelles.

Volet III : Systèmes de traduction de "troisième génération"

Dans la mesure où les volets précédents atteindront leurs objectifs, notre programme apportera à moyen terme une contribution tangible aux problèmes de la traduction. Toutefois, il faut reconnaître que cette contribution restera relativement modeste : le poste de travail n'automatise que les tâches périphériques à la traduction alors que les systèmes spécialisés ne s'attaquent qu'au cas spécial des sous-langues simples.

Si l'on veut parvenir à repousser un tant soit peu ces limites, il est absolument nécessaire de favoriser la recherche orientée vers des systèmes plus intelligents. Si les recherches sur le traitement des langues naturelles et l'intelligence artificielle ont reçu un appui considérable depuis quelques années, il demeure que ces recherches ont rarement été focalisées sur le problème de la traduction automatisée, qui constitue pourtant l'un des meilleurs bancs d'essai qui soit.

Il nous apparaît tout à fait opportun de faire maintenant l'effort d'appliquer les plus récentes techniques de l'intelligence artificielle et du traitement des langues naturelles (p. ex., différents systèmes de représentation des connaissances et de raisonnement "non-monotonique") au problème de la traduction. De plus, il faut absolument poursuivre la recherche sur les problèmes qui relèvent davantage de la traduction, comme l'acquisition et la représentation du savoir contrastif interlinguistique.

Par conséquent, notre programme prévoit l'exploration de modèles avancés en traduction automatisée (modèles dits de "troisième génération") et l'illustration de tels modèles dans des systèmes prototypes.

## CONCLUSION

La traduction met en jeu les capacités cognitives globales de l'humain. Pour cette raison, la machine à traduire universelle demeure présentement hors d'atteinte. Il est quand même possible de venir en aide aux traducteurs, dont le fardeau ne cesse d'augmenter, pour autant que l'on adopte des stratégies de sous-optimisation adéquates (niveau d'automatisation moins élevé, ou spécialisation des systèmes). Face au regain d'intérêt à l'échelle mondiale pour l'automatisation de la traduction, il devient pressant d'agir.

Le programme du CCRIT vise à redonner au Canada son rôle traditionnel de chef de file en la matière. À court et à moyen terme, ce programme adopte une approche réaliste, fondée sur des stratégies claires de sous-optimisation : poste de travail spécialisé en traduction, et systèmes orientés vers des sous-langues. À plus long terme, il vise à mettre l'intelligence artificielle au service de la traduction.

NOTES

1. Bar-Hillel Y., The Present Status of Automatic Translation of Languages, in F.L. Alt. (ed.), Advances in Computers, vol. 1, Academic Press, New York, 1960, pp. 91-163.

2. Kay M., The Proper Place of Man and Machines in Translation, AJCL, 1976, microfiche 46.

3. Kittredge R., Lehrberger J. (eds.), Sublanguage: Studies of Language in Restricted Semantic Domains, De Gruyter, New York, 1982.

4. Isabelle, P., "The TAUM machine translation project", La Revue québécoise de Linguistique, 5:2, 1986.

5. ALPAC. Languages and Machines: Computers in Translation and Linguistics. Report of the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, NCR, publication 1416, Washington, 1966.

6. Machine Translation and Natural-Language Processing: Opportunities for Artificial Intelligence in Canada. Report of a study by the firm Cognos under contract from the Department of Communications, ref: DOC-CR-84-048, 1985.